Applications of computers & optical character recognition devices

J. H. Perry Jr. (read by C. J. Duncan)

This paper gives the history of Perry's automation program for almost four years, describes equipment used (including that developed specially by and for Perry Publications), indicates savings over previous methods and projects future developments in the industry.

Included is such equipment as the Retina Reader, which reads printed or typewritten copy up to a rate of 2,400 characters per second or approximately 28,000 average-length words per minute; the Perry Photo Composer which could bring far-reaching changes in cold-type ad and newspage production; the integrated RCA 301 and Photon systems; book, catalogue and encyclopedia production, etc.

The Perry organization, with executive offices in West Palm Beach, Florida, publishes 27 Florida newspapers; the *Freeport News* in the Bahamas, *Palm Beach Life Magazine* and *All-Florida Magazine*. The state-wide All-Florida News Service is also operated by Perry Publications. While primarily in the newspaper business, Perry also is a major producer of commercial printing. Although Perry's primary concern is with newspapers, unforeseen capabilities of the computer have enabled them to lease time for outside data processing and to embark on book production. So far, the computer equipment has been used to produce more than 3,000 books.

Use of electronic computers in printing and publishing began only three years ago, when the Radio Corporation of America suggested Perry use a computer not only for accounting, as they had been doing for two years with IBM equipment, but also for preparation of typesetting. Both the offset and letterpress processes are used; and both hot metal and cold type. Early expectation was that automation would enable greater use of less-costly cold type. The expectation has come true. The first computer was the RCA 301. Its primary purposes were bookkeeping, hyphenation and justification. Plans for installing a magnetic drum as a storage medium for the dictionary and the various programs now in use have been changed temporarily. Instead of a magnetic drum, Perry installed a data disc file, which has a storage capacity of 88 million characters.

One feature of this device is the ability to store in it an 'exception word dictionary' which is utilized by all composition programs. The dictionary is updated with all words once found to be erroneously hyphenated. As a result, all stations serviced by the computer centre receive equal consideration by providing the proper names peculiar to their particular area or subject. The data disc file should not significantly affect the throughput speeds of any of Perry's composition routines. The idea of eventually installing the magnetic drum as the primary storage medium has not been abandoned. A final decision has been deferred until the forthcoming delivery and installation of the new computer generation complex, the RCA Spectra 70.

Another piece of computer hardware has been installed on the supporting system. It is an RCA paper tape reader with 1,000 cps capabilities and a paper tape punch with 100 cps output speed. This replaced an original RCA combination paper tape reader-punch which had a transfer rate of 100 cps. The throughput speed of this system is approximately 80 cps for all composition routines.

As for hot metal straight-matter composition, modifications are made from time to time, but the program is basically the same as that in use since December 1962. This program currently is capable of handling any one of the 20 available fonts contained in the tables. Any font can be added to this program within a 24-hour period. Photo typesetting also is used on the ATF-CS.

The program operates in the same manner and is capable of producing the same data as the hot metal program for any one of the five fonts available on discs. Perry installed a number of the Model CS at smaller Perry papers. Among the attractive features of the Model CS are its small size, its relatively low cost, ease of maintenance and its consistent 10 lines per minute production. While the ATF Model CS satisfies the immediate needs of the smaller papers, it is not capable of meeting demands for high-speed computerized production in larger plants. The answer to this problem appears to be the Photon 713, which is capable of 35 lines of straight-matter composition per minute. Perry has received delivery of one 713 and five others are on tentative order. Since the 713 is primarily a straight-matter machine, it will not replace the Photon 513s now in operation. Current program for the 513 covers 15 different type styles and 90 special characters (pi) that can be delivered in any one of twelve sizes ranging from 6 to 72 pt in variable line length up to 42 picas. Another new development in conjunction with the Photon 513 is what Perry calls a commercial disc. It is especially designed for composition of books, magazines, brochures and similar matter.

Page makeup

During recent months Perry has had a sort of three-way wedding of production facilities in the West Palm Beach plant. It involves the computer, the Photon 513 and the Perry Photo Composer, an automatic full-page makeup machine.

Like the Photon 513, the Photo Composer takes punched tape from one of the 301s and this tape instructs its own computer to position and enlarge the film from the 513 to complete the automatic page makeup. The machine is capable of automatically putting on borders, but does not add illustrations and artwork. These two items are stripped in by artists in the quality control department. At present, two programs are in production for the composer.

The first is block copy. This block copy, consisting of any mix required, can be delivered in a single exposure on the composer, providing it is not greater than 30 picas wide and 4″ deep. The second is line copy. This line copy, consisting of any one parameter, can be delivered on the composer for any width up to a page, and any size, depending on the size enlargement requested, up to an actual limit of 8 times 72, or 576 points. Use of the Photo Composer for newspage makeup also is being experimented with.

Optical character recognition

The full name of the equipment is the Electronic Retina Character Reader system. It is built by Recognition Equipment Inc., of Dallas, Texas, and Electronic Retina is the trademark of that firm. This new equipment is compatible with and has been integrated into our RCA 301 computer facilities. The Electronic Retina accepts typewritten matter as an input medium to a computer rather than conventional unjustified paper tape.

The system consists of: A rapid index page carrier, an Electronic Retina and recognition unit, a Scientific Data system 910 computer, and two Ampex magnetic tape stations which are compatible with RCA. Hard copy is fed into the system, and the characters are read by the Electronic Retina. This hard copy, prepared with special typewriters developed for us by Olivetti, can be editorial text matter, classified advertising set solid, semi-display, display, legals, accounting data, or any other alphanumeric information. The Electronic Retina converts the typewritten characters and spaces into predetermined electrical impulses which are stored on magnetic tape. The magnetic tape, in turn, produces justified paper tape to drive the linecasting machines—either hot or cold type. Printed or typewritten copy is read up to the rate of 2,400 cps (approximately 28,000 words per minute).

The machine processes pages which vary in size from $3\frac{1}{4}'' \times 4\frac{7}{8}''$ to 14″ by 14″ and in weight from 12 lbs paper to 30 lbs paper at a rate which varies from 10 to 30 pages per minute, depending upon total information to be read from any page. At any one time, different sized pages, within the range just given, can be handled intermixed as long as there is no more than four inches difference in length among the pages in the batch. The machine closely resembles the functioning of a human eye, and views constantly rather than intermittently. It senses black, white, and all shades of grey. It also adds a slight amount of 'jitter' when viewing a subject, to smooth out the edges of rough lines.

The Electronic Retina is capable of the following tasks:

producing idiot magnetic tape for the 301 to set all local display advertising, using the present 301 programs
producing idiot magnetic tape to set all classified display and set solid classified advertising, using the present 301 programs
producing idiot magnetic tape to handle all accounting, using the present 301 programs
reading and transmitting idiot magnetic tape to the 301 for all text matter programs. Magnetic tape stations are compatible to the 301 tape stations. As a result, it is possible to go from magnetic tape to magnetic tape by taking by hand magnetic tape from the Electronic Retina and positioning it on the 301 tape station by hand. After typed copy is read by proofreaders and editors, the copy is corrected before reading by the Electronic Retina by using the correction routine.

Corrections

Corrections are made before composition is accomplished. Or, rather, making the corrections before the Electronic Reader provides the idiot tape for the 301. Each line on each typewritten page of hard copy is identified by number. Editing and proofreading changes are marked on the hard copy before it is inserted in the Electronic Retina. These changes and corrections are marked in the conventional manner on regular copy and, in addition, an identifying mark (usually numerical sequence) is made in the right hand margin at the end of the affected line.

When the editors and proofreaders complete their work on the copy, the original is returned to a typist at one of the special typewriters.

The typist then retypes the lines, or segments of lines, affected by the changes, using a new piece of copypaper as a correction sheet. The correction sheet then is fed into the Electronic Retina ahead of the original copy which is to be corrected.

The reader stores these corrections and its coded instructions concerning them in its memory system. Then, as the regular text comes in behind, the correction is completed. Under this routine, the Retina will accept corrections ranging from transposition of two letters in a single four-letter word to insertion of dropped copy ranging up to an indefinite number of lines. Then it returns to the regular text, accepting it without interruption until another corrected line is encountered.

Programmers have been at work on an advertising insertion order for national, local, classified, legal and accounting requirements. The insertion order contains the code number for accounting production, classification, and the number of the product, as well as all other insertion instructions and other information.

It also enables Perry to transmit the exact information from the advertising insertion order room and actual account. If it sets solid classified, it sends the composition to the composing room and it counts the number of lines of copy so that we can have the exact line count for makeup purposes. When the last ad is stored in the computer, Perry Publications know exactly how much space to allot for the classified advertising section.

At the end of 1966 Perry expected to receive delivery of an RCA Spectra 70 computer complex. The Spectra 70, in basic design, is a total management information system to meet our needs now and for many years in the future. This it can do on-line and in real time. It works on what is known as monolithic integrated circuit. It harmonizes with most computers, it will save the present programming investment and will eliminate costly reprogramming, retraining and reinvestment. Its speed is measured in nano-seconds: This will be an important step ahead in the crucial dictionary retrieval system.

Perry Publications have been working on the cold type process continuously since 1946. The only bottleneck is the need for a high-speed straight-matter machine. A solution seems near, either by installation of several Photon 713s with their 35 lines per minute speed, the Photon 900 with its 150 lines per minute speed, or an electron gun or page generator with speeds in the neighbourhood of a full, complete page in less than 200 seconds.

Another vital function which the Spectra 70 can perform as a total management information system is to provide a business profile analysis. Perry will program into the memory facilities a wide range of pertinent information concerning each Perry property. This would include not only data regarding actual operations of the property, but significant details about the specific community and area, such as population, growth trends, degree of circulation penetration by households, employment, industrial expansion, degree of competition from other media, and so on.

Current operating data will be combined with this basic stored-up information at regularly designated intervals (probably every seven days—and management will be provided with a detailed performance report far more elaborate and informative than the conventional operating statements in use today). These reports would not only indicate trouble areas immediately after they develop, but would show trends and indications that would enable management to head off some trouble spots before they could develop.

So far as the newspaper's future is concerned, it is hoped one day to use the computer to evaluate and qualify news in order to provide the editorial content of the papers with proper balance. An editor could quickly review the news story from the printed copy and assign each story a category which could be encoded for future computer use. Additions or corrections could also be encoded and because a computer can digest and arrange twelve hours of normal copy in roughly 15 minutes, the editor is freed of the humdrum task of sorting his own news. All original news stories would be translated from the paper tape on to a magnetic tape which would then act as a memory. A second tape would be prepared, listing all corrections, additions and deletions. Each news item would be prefaced by a category number and a priority rating. A third tape contains the news balance formula the editor is using as well as the size of the 'news hole'. Inside the computer, all news could be assigned to categories, such as: government and politics, war, defence, diplomacy; economics, business and travel; crime; public moral problems; health and welfare; accidents and disasters; science and invention; education, classic arts, religion; popular amusements, mass media; and general human interest.

After the size of the 'news holes', emphasis for each category and allowable minimum length for each story have been determined, here's what would happen:

category percentages are multiplied by the news hole figure to determine desired lines per category
memory tapes containing all available news are searched; lines are counted and selected
desired lines per category are revised to reflect surplus or deficiency
desired lines per story are set according to editing formula and story priority
available story lines are matched with desired amounts
results are revised to reflect surplus or deficiency in any particular category
high-speed printer produces copy arranged according to categories with each line one news column width.

One advantage of such use of the computer would be to fortify the newspaper against charges of being unfair. Almost everybody at some time has felt that the value content of a news item was unbalanced—and frequently this has been so. The computer, if properly applied, could prevent any future injustice.